

文章编号:1001-9081(2015)08-2311-05

doi:10.11772/j.issn.1001-9081.2015.08.2311

基于道路环境上下文的行人跟踪方法

方义¹,嵇智源^{2*},盛浩³

(1. 中航发动机控股有限公司 科技与信息化部, 北京 101304;

2. 中华人民共和国科学技术部 高技术研究发展中心, 北京 100044; 3. 北京航空航天大学 计算机学院, 北京 100191)

(* 通信作者电子邮箱 jzy@htrdc.com)

摘要:针对目前城市交通中人车混行场景中行人跟踪效果不佳的问题,提出了一种基于道路环境上下文的行人跟踪方法。首先通过对道路环境上下文进行分析,建立道路模型;其次在道路模型的约束下建立行人与环境的交互运动模型;最后利用该模型进行行人的跟踪。在真实场景中的实验表明使用了道路上下文信息的跟踪方法与连续离散连续能量最小化的多行人跟踪方法相比,多目标跟踪准确度从 47.6% 提升至 63.2%,多目标跟踪精度从 68.8% 提升至 74.3%。数值结果表明道路上下文信息对于提高人车混行场景中行人跟踪效果的有效性。

关键词:多行人跟踪;人车混行;环境上下文;道路模型;环境交互

中图分类号: TP391.4 **文献标志码:**A

Pedestrian tracking method based on road environment context

FANG Yi¹, JI Zhiyuan^{2*}, SHENG Hao³

(1. Department of Technology and Information, Aviation Industry Corporation of China, Beijing 101304, China;

2. High Technology Research and Development Center, Ministry of Science and Technology of the People's Republic of China, Beijing 100044, China;

3. School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: Since tracking pedestrians with car driving is still a problem in multi-object tracking, a pedestrian tracking method based on road environment context was proposed. Firstly, with the analysis of road environment context, an interaction algorithm based on the path model was proposed for pedestrians' motion prediction, then the model of pedestrians mixed with vehicles was introduced. Finally, the model was applied in pedestrian tracking. The experimental results show that compared with discrete-continuous tracking approach, Multiple Object Tracking Accuracy (MOTA) of the proposed algorithm grows from 47.6% to 63.2% and Multiple Object Tracking Precision (MOTP) grows from 68.8% to 74.3%. The results prove the effectiveness of road environment context to improve the pedestrian tracking effect in mixed vehicle scene.

Key words: multiple pedestrians tracking; pedestrians mixed with vehicles; environment context; path model; interaction with environment

0 引言

近年来,世界城市化进程飞速发展,在过去的十年里,中国的城镇化水平从 36% 增长到 51.27%^[1],城镇化对于现在社会的经济和人民生活水平提高都有很大的好处,但是随着城镇化进程的加快,同样也带来一系列问题^[2],如交通拥堵、环境污染、交通安全等。在中国,交通拥堵是一个非常严重的问题,据统计现在北京的平均车速是 20 km/h,在一些比较拥挤的城市交通路口,甚至低到 7~8 km/h 的水平^[3];并且城市基础设施并不完善,如人车不分流,存在很多人车混行的场景^[4],导致城市运行效率低。对于该环境下行人进行跟踪,对于智慧城市的建设具有非常重要的意义。

本文探索并分析了人车混行场景下的行人在该场景下的行为模式,并提出了面向人车混行场景的行人跟踪方法。

1 相关工作

多目标跟踪技术作为一个活跃的研究领域已经发展了几

十年。在 20 世纪七八十年代,多目标跟踪理论得到了迅速的发展,许多经典模型如简单的时间相关模型、概率数据关联模型、联合概率数据关联模型^[5]、多假设跟踪^[6]、超图匹配^[7]等算法纷纷被提出^[8]。近年来,粒子滤波器^[9]、基于马尔可夫链蒙特卡洛方法的数据关联算法^[10]、目标行为分析^[11]、分层融合^[12]、能量最小化^[13]等新技术的出现给多目标跟踪技术的进一步发展带来了新的活力和研究方向。多目标跟踪的核心为数据关联技术,如联合概率数据关联滤波器是一种较为常用的数据关联方法,它能较好地利用目标运动模型和图像观测信息进行数据融合和状态更新,但缺乏有效手段处理新进入视场、离开视场和被遮挡的目标^[8]。多假设跟踪通过记录和比对多个全局假设,能够克服联合概率数据关联滤波器的这个缺点,其将关联匹配的决定进行适当延后,直到未来的某个时刻有较为明晰的线索出现,再进行优化抉择^[6]。针对复杂场景中的跟踪采用层次化的轨迹关联算法^[13],能够对不同情况的轨迹关联状态进行分层处理,适应复杂的轨迹关联

收稿日期:2015-01-12;修回日期:2015-04-24。

基金项目:国家自然科学基金资助项目(61472019);国家 863 计划项目(2013AA01A603)。

作者简介:方义(1974-),男,安徽池州人,高级工程师,博士,主要研究方向:模式识别、信息系统集成;嵇智源(1956-),男,北京人,高级工程师,主要研究方向:信息领域科技管理、软件工程管理;盛浩(1981-),男,浙江台州人,副教授,博士,CCF 会员,主要研究方向:计算机视觉、模式识别。

过程求解。

多目标跟踪技术中的另一个关键问题是如何准确预测目标运动状态及可能出现的位置。早期的研究仅利用速度及加速度动力学方程模拟目标的运动情况,结合卡尔曼滤波、粒子滤波等方法降低检测噪声以实现对目标的位置预测^[6]。然而,这些方法普遍忽视了目标间的互动关系以及环境对目标的约束关系,对于人车混行场景,其位置预测精度不够,给后续的多目标关联造成巨大挑战。而在交通流建模领域,通过研究目标的运动行为及互动关系,实现了多目标运动效果^[14-15]。通过上述研究可知,目标的运动模式并不仅仅是简单的线性模型,可能会受到各种制约而产生非线性的运动模式^[10,16]。对于人车混行的城市环境,通过在线学习,能够快速更新目标的运动模式,并能对目标的运动情况、目标间的相互作用关系,以及环境对目标的约束关系进行较为准确的描述。

2 跟踪框架

本文使用一个预先训练好的 Deformable Parts Model (DPM)^[20]行人检测器^[21]对一段视频序列中的每一帧进行目标检测,得到检测结果 $D = \{d_1^t, d_2^t, \dots, d_i^t\}$ 。其中 $d_i^t = \{x_i, y_i, s_i\}$ 代表目标 i 在时刻 t 的位置 (x_i, y_i) 和大小 s_i 。首先,使用卡尔曼滤波对目标进行关联,得到短轨迹 $T_L = \{T_L^1, T_L^2, \dots, T_L^n\}$, 设在关联之后生成的轨迹集合为 $T_M = \{S_1, S_2, \dots, S_n\}$, 其中 $S_k = \{T_{i_1}^L, T_{i_2}^L, \dots, T_{i_k}^L\}$ 是一条关联轨迹中的短轨迹集合。这个关联过程可以转换为一个最大化后验概率问题:

$$\begin{aligned} s^* &= \arg \max_{S_k} P(S_k | T_L) = \\ &\arg \max_{S_k} P(T_L | S_k) P(S_k) = \\ &\arg \max_{S_k} \prod_{T_L^i \in T_L} P(T_L^i | S_k) \prod_{T_L^i \in S_k} P(T_L^i) \end{aligned} \quad (1)$$

对于这个问题的求解,可以采用匈牙利算法直接求解。可以将该最大化后验概率问题转换为二分图匹配的问题,采用式(2)的权重矩阵。为了得到完整的轨迹,这里采用多次迭代式(1)的方法得到完整轨迹,在每次的迭代中增加轨迹之间的间隔。

$$\mathbf{M}_{\text{cost}} = \left(\begin{array}{c|c} \mathbf{M}_{\text{link}} & \mathbf{M}_{\text{init}} \\ \hline \mathbf{M}_{\text{term}} & \mathbf{0} \end{array} \right)_{2n \times 2n} \quad (2)$$

其中: \mathbf{M}_{init} 是一个对角矩阵,每个对角元素代表当前轨迹片段作为关联后轨迹片段起始片段的概率; \mathbf{M}_{term} 也是一个对角矩阵,每个对角元素代表当前轨迹片段作为关联后轨迹片段起始片段的概率; \mathbf{M}_{link} 是两个轨迹片段关联的概率子矩阵,每个元素代表两个轨迹片段相关联的概率,如式(3)所示,由轨迹片段间的外观相似度、运动相似度、时间相似度三部分决定。

$$\mathbf{M}_{\text{link}}(T_L^i, T_L^j) = p_a(T_L^i, T_L^j) \cdot p_m(T_L^i, T_L^j) \cdot p_t(T_L^i, T_L^j) \quad (3)$$

3 基于环境上下文的行人跟踪方法

在交通场景中,行人的行为很大程度上受到道路交通场景以及交通规则的约束。通过对视频序列中行人活动的观测值建立道路模型,利用道路环境上下文完成行人的运动预测。

3.1 道路模型应用于多行人跟踪

视频中行人或者车辆的共同行为模式反映到视频图像序列中即为视频中像素运动具有相同的规律,因此可将视频中高级行为模式转换为低层次的像素运动规律发现问题。通过对视频像素变化的测量,计算其光流信息,然后通过对光流信息的分析,可以得到目标原子行为,通过原子行为的分析完成道路模型建立过程。本文使用文献[20]中提出的方法对道路模型进行发现,如图 1 所示。

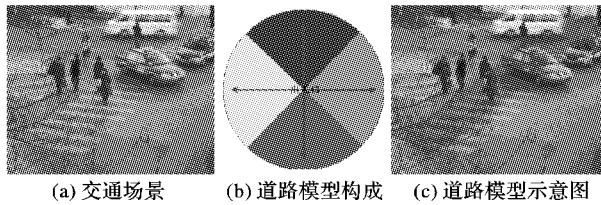


图 1 光流量化方向与道路模型建立

在获得道路模型之后,将道路模型应用于目标间的轨迹关联,其可直接用于计算轨迹起始和终止概率。通常情况下,如轨迹起始点靠近道路出入口,则该条轨迹为起始轨迹的可能性就越大;如轨迹终点靠近道路模型出入口,则该轨迹为终止轨迹的可能性就越大。因此,定义轨迹的起始概率与终止概率计算如下:

$$P_{\text{init}} = \exp(-\|p_{T_i^M}^{\text{tail}} - V_d\| / C_d) \quad (4)$$

$$P_{\text{end}} = \exp(-\|p_{T_i^M}^{\text{head}} - V_o\| / C_d) \quad (5)$$

其中: $p_{T_i^M}^{\text{head}}$ 是轨迹 T_i 的起始位置, $p_{T_i^M}^{\text{tail}}$ 是轨迹 T_i 的结束位置, V_d 和 V_o 分别是当前轨迹距离最近的出入口位置。城市道路通常为双向道路,因此在本研究中出口与入口的选择是相同的,如图 2 所示。



图 2 道路模型对轨迹起始终止概率的影响

3.2 行人与环境交互

行人在人车混行场景下,运动模式与其在普通场景下会有较大不同,其行为通常会受其他行人或者车辆等外在因素的影响而导致不确定性增大,同时其运动受道路环境的严格约束,如图 3 所示。这里为了刻画行人在人车混行下的运动模型,采用基于能量最小化的方法对行人的运动进行建模。

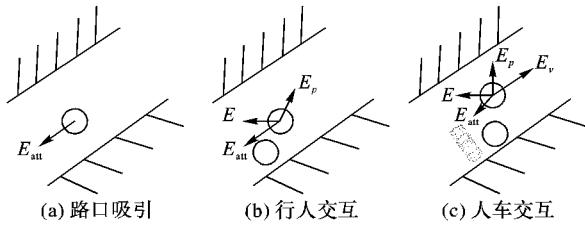


图 3 行人与环境交互示意图

为了建立精确的行人运动模型并减少模型的复杂度,进行快速计算,本研究只考虑对行人的运动模型影响最大的三

个因素:1)目的地对行人的吸引力 $E_{\text{att}}^i(p_i, v_i)$;2)行人间的排斥力 $E_{pp}^i(p_i, v_i)$;3)车辆对行人的排斥力 $E_{pe}^i(p_i, v_i)$ 。对以上三个部分进行加权得到能量函数:

$$E(p_i, v_i) = E_{\text{att}}^i(p_i, v_i) + \lambda_1 E_{pp}^i(p_i, v_i) + \lambda_2 E_{pe}^i(p_i, v_i) \quad (6)$$

3.2.1 目的地对行人的吸引

由于道路的出入口是固定的,通过行人当前的运动方向判断出行人将要到达的目的的出口方向,找到距离行人当前位置 p_i 最近的出口 O_i ,行人当前速度 v_i 和出口 O_i 之间的差异代表其作用力:

$$E_{\text{att}}^i(p_i, v_i) = \frac{O_i - p_i}{\|O_i - p_i\|} \cdot \frac{v_i}{\|v_i\|} \quad (7)$$

3.2.2 人人交互模型

为了对行人之间的交互模型进行仿真,假设已知每个行人的运动状态,并能预测下一时刻每个行人的位置及与其他行人之间的位置关系。人人交互因素主要包括行人之间的距离 σ 与运动方向夹角 φ ,通过不同的参数取值,将人人交互模型分为群组和离散两种关系。当 σ 小于阈值 ε 且 φ 小于 θ 时,行人之间的关系归为群组关系。根据实验经验值, ε 取跟踪框的宽度, θ 取 $\pm 5^\circ$ 。在本文方法中,对于测试集中的每一段视频序列,取前 25% 帧作为训练视频序列,后 75% 帧作为测试序列。通过训练得到人与人交互模型的各参数值。设当前目标 s_i 具有速度 v_i ,行人通常与其他行人保持一定距离 σ_d ,定义行人 i 与其他行人 j 间的排斥力为:

$$E_{ij}^i(v_i) = \begin{cases} \exp(-d_{ij}^2(v_i)/2\sigma_d^2), & i \text{ 与 } j \text{ 为离散关系} \\ 0, & i \text{ 与 } j \text{ 为群组关系} \end{cases} \quad (8)$$

由于道路中的多位行人对同一行人的作用会叠加,并且与两个行人之间运动方向的夹角有关,因此行人间的作用能量为:

$$E_{pp}^i(p_i, v_i) = \sum_{k \neq i} w_{ij}^e(S_i, S_j) \cdot E_{ij}^i(v_i) = \begin{cases} \sum_{k \neq i} ((1 + \cos(\varphi_{ij})/2) \cdot E_{ij}^i(v_i)), & i \text{ 与 } j \text{ 为离散关系} \\ 0, & i \text{ 与 } j \text{ 为群组关系} \end{cases} \quad (9)$$

其中 φ_{ij} 为行人 i 与行人 j 运动方向之间的夹角。

3.2.3 人车交互模型

在人车混行场景下,车辆对于行人影响往往大于行人间的影响,因此,行人对于车辆的反映比对行人的反映更加强烈,体现在能量函数中为能量值更大。车辆对于行人影响表现为当车辆在距离行人距离比较远的情况下,行人反映比较平静;当车辆对于行人较近时,行人运动变化剧烈。对于这种运动模式本研究通过 Sigmoid 函数来表示:

$$E_s^i(v_i; S_i, S_j) = 1 - \frac{1}{1 + \exp(-[d(S_i, S_j) - \sigma_m]/\sigma_v)} \quad (10)$$

其中: σ_m 表示行人 i 与车辆 j 之间期望保持安全距离的阈值,一旦小于这个阈值则行人的运动可能会发生剧烈变化,大于这个阈值可能不受影响; σ_v 表示行人行为发生剧烈变化随距离变化的变化率。在考虑车辆运动的情况下,还跟行人与车辆运动方向的夹角有关,车辆 j 对于行人 i 的能量为:

$$E_{pe}^j(v_i; S_i, S_j) = w_i(S_i, S_j) E_s^i(v_i; S_i, S_j) =$$

$$((1 + \cos(\varphi_{ij} - 2^{-1}\pi))/2) E_s^i(v_i; S_i, S_j) \quad (11)$$

场景中车辆对行人的影响并不取决于所有车辆影响的叠加,而是取决于所有车辆中对行人影响最大的车辆,所以:

$$E_{pe}^i = \max_j \{E_{pe}^j(v_i; S_i, S_j)\} \quad (12)$$

3.2.4 能量函数应用于跟踪

对于当前短轨迹 T_L ,其与下一时刻其他短轨迹关联的运动概率如下:

$$\begin{aligned} p_{\text{motion}}(T_i^L | T_j^L) &= \\ G(p_{\text{tail}_i} + \tilde{v}_i \Delta t; p_{\text{head}_j}, \sum \Delta t) \cdot \\ G(p_{\text{head}_j} - \tilde{v}_j \Delta t; p_{\text{tail}_i}, \sum \Delta t) \end{aligned} \quad (13)$$

其中 \tilde{v} 为使用能量函数预测的当前运动速度。式(6)中的五个参数确定之后才能在轨迹关联过程中使用能量函数并进行最小化求解最优速度。在能量函数中参数 $\sigma_d, \sigma_m, \sigma_v$ 为行人运动属性,其值与场景相关。对于同一场景,参数取值相同;对于不同场景,参数取值并不相同。本文采用统计方式对这些参数赋值,图 4 是对实验数据集的一段路口视频(场景如图 2 所示)的行人与车辆标注结果进行统计得到的部分属性统计图,其中, $\sigma_d = 43, \sigma_m = 190, \sigma_v = 25$ 。从图中可以看出,行人在遇到车辆或者行人之后会进行突然的行为改变,但是整体上行人的运动还是呈现一定的运动规律,行人的速度、行人与行人之间的间距、行人与车辆之间的间距基本都服从高斯分布,行人跟车辆的距离对行人速度的改变呈符合 Sigmoid 函数趋势。使用标注的视频数据对每一项能量进行求解,两个目标的关联问题可看作二分类问题,这里使用最简单的最小二乘法进行求解完成对所有参数的计算。

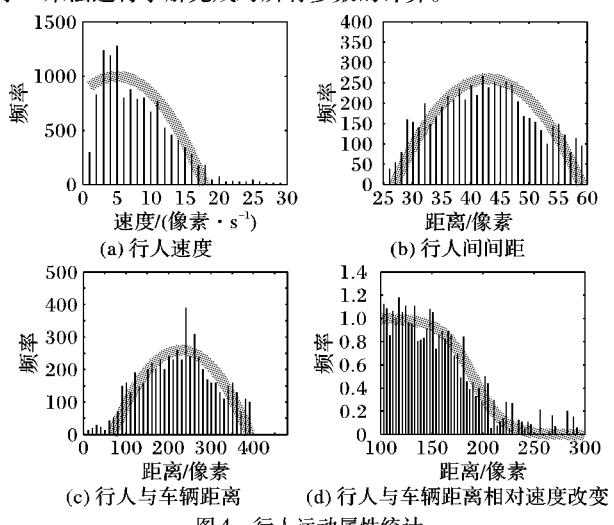


图 4 行人运动属性统计

4 实验验证

为了验证人车混行场景下基于环境上下文的行人跟踪方法的有效性,数据集采用了真实的人车混行场景视频,由架设在公路斜上方的摄像机拍摄采集,视频分辨率为 980×780 像素,数据集包括交汇路口(丁字、十字)、混行道路(狭窄、宽阔)、单行道等 25 个视频片段,覆盖了行人间相互遮挡、行人运动行为突然改变、行人受车辆遮挡等情况。

为了验证本文方法的有效性,结合 Li 等^[20]提出的评价指标与 Stiefelhagen 等^[21]提出的 CLEAR MOT 指标对本文的行人跟踪算法进行评价,主要评价指标如下:

- 1) 多目标跟踪准确度 (Multiple Object Tracking Accuracy, MOTA)。
- 2) 多目标跟踪精度 (Multiple Object Tracking Precision, MOTP)。
- 3) 目标 ID 变化数 (ID Switch, IDS)。被跟踪目标 ID 跟标注数据对比 ID 变化的次数。
- 4) 轨迹片段数 (FragMent, FM)。跟踪结果中, 轨迹断裂的次数。

4.1 不同特征对跟踪结果的影响

为了验证道路模型、人车模型和人人模型对跟踪结果的影响, 本实验仅对人车混行场景下的行人进行跟踪, 以短轨迹关联结果作为基准, 分别添加道路模型、人与环境上下文的交互模型来验证基于道路环境上下文的行人跟踪方法的有效性。

4.1.1 道路模型对轨迹起止概率计算的影响

道路模型主要对目标的起始轨迹与终止轨迹的计算进行影响, 对于短轨迹关联中, 单独加入道路模型来判断道路模型的作用, 是长轨迹的起始与终止位置的分布在实验视频序列上的展示。图 5(a)是未加入道路模型之前轨迹起始点与终止点的分布, 图 5(b)是加入之后的分布。从图中可以看出在加入道路模型之后分布在道路中间的轨迹的起始和终止位置明显减少。



图 5 道路模型对起始终止轨迹概率计算的影响

4.1.2 交互模型对运动预测的影响

为了验证交互模型对行人位置预测的准确性, 本实验中使用视频序列的标注数据进行位置的预测, 并且与线性运动模型方法以及卡尔曼滤波方法进行了对比实验。如图 6 所示, 横坐标表示预测的位置与真实位置距离的阈值。实验结果可以看出, 随着阈值的增大, 三种方法位置预测的准确性都增加, 但是交互模型的预测准确性增长速度明显大于线性运动模型与卡尔曼滤波模型。当阈值增大到 70 像素距离时, 交互模型预测的准确性要明显优于线性模型和卡尔曼滤波。

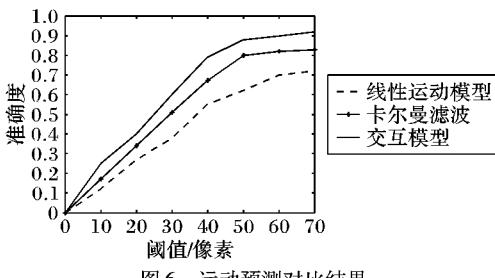


图 6 运动预测对比结果

4.1.3 不同特征对跟踪结果的影响

在轨迹关联过程中, 分别加入道路模型和交互模型, 计算两者对于跟踪结果的影响, 实验结果如表 1 所示。从中可以看出, 添加了道路模型之后, MOTA 有所提高, 并且 IDS 和 FM 都有所减少, FM 减少主要是因为道路边缘的轨迹由于道路模型的加入, 使得这部分轨迹作为长轨迹的起始或者终止的

概率计算更加准确, 因此使得 IDS 和 FM 减少。在增加了人人与人车交互模型之后, 对于行人的行为预测更加准确, 能够解决行人在运动过程中由于与其他行人或车辆交互导致的行人运动突变。由于知道了车辆的运动信息, 可预测行人在车辆遮挡情况下的运动状态, 从而进行正确的轨迹关联, 因此交互模型使得 MOTA 和 MOTP 都有显著提高, 并且大幅减少了轨迹片段数目和 ID switch 的数目。人人交互模型中群组信息的加入, 能够有效地解决群组内部发生 ID switch 和跟踪失效引起的轨迹断裂等情况, 使得本文跟踪算法对于群组内行人之间的抗遮挡能力显著提升。如图 7 所示, 传统的连续离散能量最小化的多行人跟踪方法 DCT (Discrete-Continuous Tracker) 与加入人人交互模型的 DCT 的效果进行比较, 后者相对于前者有显著提升。因为两个模型在计算轨迹关联时会有一定的关联结果重合, 当两个模型同时作用后, 两个模型的效果并没有叠加。

表 1 使用不同的特征对跟踪效果的影响

模型	MOTA/%	MOTP/%	IDS	FM
基准	49.6	65.9	110	63
道路模型	56.3	66.2	96	50
人车交互模型	60.1	73.1	90	23
人人交互模型	67.7	75.3	36	25
道路模型 + 交互模型	73.2	77.8	29	10

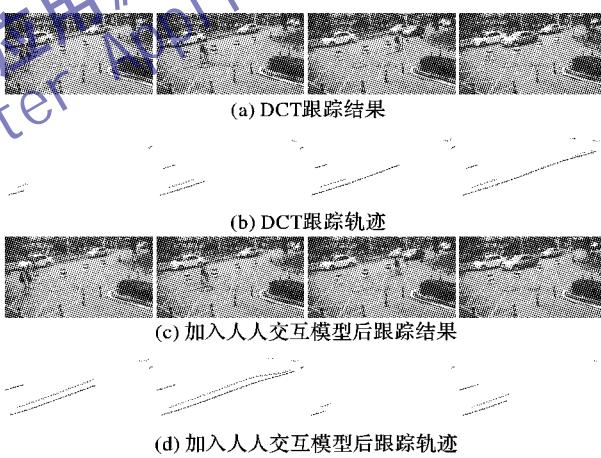


图 7 人人交互模型对实验结果的影响

4.1.4 不同跟踪方法的实验对比

针对人车混行场景, 利用道路模型和人人与人车交互模型在一定程度上解决了人车混行场景下的多行人跟踪问题。为了证明本文方法的优越性, 将本文方法与连续离散连续能量最小化的多行人跟踪方法 DCT^[17] 的实验结果进行了对比, 如表 2 所示。本文算法在跟踪的准确性上远高于 DCT 算法, 并且 FM 数目和 IDS 的数目也远少于 DCT 算法。这是因为 DCT 算法在进行轨迹关联的过程中, 仅仅考虑了不同轨迹之间的排斥作用, 没有考虑轨迹之间的交互, 并且其算法没有考虑人车混行场景的特殊性和人人交互模型中的群组关系。

表 2 使用不同方法进行跟踪效果对比

方法	MOTA/%	MOTP/%	IDS	FM	每帧平均运行时间/s
DCT	47.6	68.8	51	110	1.404
本文方法	73.2	77.8	29	10	1.263

图 8 为本文方法与 DCT 方法的跟踪结果, 由图 8(a)、

(b)可以看出对于车辆长期遮挡的两个行人,本文方法在长期遮挡之后,仍然能够对其进行正确的轨迹关联,但是DCT算法则是改变了两个目标的ID,并没有很好地进行关联。在图8(c)、(d)中行人在车辆经过时进行了速度改变,其中一个行人被遮挡,并且行人间发生了交叉,在这种情况下,DCT算

法发生了漏跟与ID改变的情况,但是本文中的算法仍然能获得良好的跟踪效果,并没有发生漏跟与ID改变。如表2所示,在计算时间上本文算法略优于DCT方法,主要由于本文提出的道路模型和交互模型对跟踪范围进行了约束,提高准确度的同时,降低了计算的空间和时间复杂度。

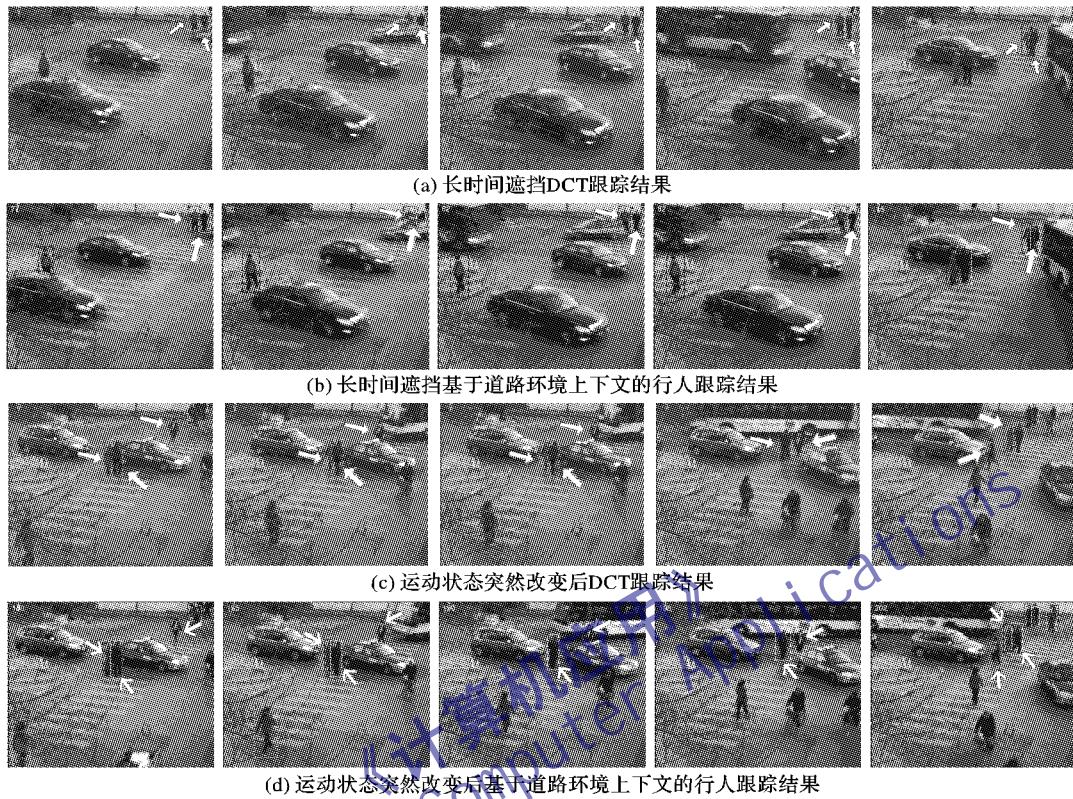


图8 跟踪实验结果对比

5 结语

本文针对行人在人车混行场景下运动状态复杂、遮挡严重等情况,提出了基于环境上下文的多行人跟踪方法。通过引入道路模型确定轨迹片段在长轨迹中的起止概率提高轨迹概率计算准确性;而后,通过行人交互模型与人车交互模型对行人的运动规律进行建模,通过能量最小化方法求解行人的运动状态,对轨迹进行正确的关联;最后,在上述理论的基础之上进行了实验验证,结果表明该方法对于人车混行场景下的多行人跟踪效果显著,跟踪结果明显优于对比算法,同时也证明了人车交互模型对于实际场景中跟踪有较大价值,后续工作可以引入更多的交互模型,如人与场景的交互、车与场景的交互等。

参考文献:

- [1] National Bureau of Statistics of China. China's total population and structural changes in 2011 [EB/OL]. (2012-01-20) [2014-11-13]. http://www.stats.gov.cn/english/pressrelease/201201/t20120120_72112.html.
- [2] ALUSI A, ECCLES R G, EDMONDSON A C, et al. Sustainable cities: oxymoron or the shape of the future? Harvard Business School Organizational Behavior Unit Working Paper No. 11-062 [EB/OL]. Boston: Harvard Business School, 2011 [2015-02-09]. <http://ssrn.com/abstract=1726484> or <http://dx.doi.org/10.2139/ssrn.1726484>.
- [3] CHEN Z, FAN W, XIONG Z, et al. Visual data security and management for smart cities [J]. Frontiers of Computer Science in China, 2010, 4(3): 386–393.
- [4] GAO Y. Research and implementation of intelligent control for pedestrians and vehicles mixed environment [D]. Chengdu: University of Electronic Science and Technology of China, 2012: 31–57. (高杨. 人车混合环境智能控制研究与实现 [D]. 成都: 电子科技大学, 2012: 31–57.)
- [5] XIAO J, CHENG H, HAN F, et al. Geo-spatial aerial video processing for scene understanding and object tracking [C]// CVPR 2008: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2008: 1–8.
- [6] COX I J, HINGORANI S L. An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(2): 138–150.
- [7] BRENDL W, AMER M, TODOROVIC S. Multiobject tracking as maximum weight independent set [C]// CVPR '11: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2011: 1273–1280.
- [8] WANG X, ZHAO R. Person re-identification: system design and evaluation overview [M]// Person Re-Identification. Berlin: Springer, 2014: 351–370.
- [9] GONCZAREK A, TOMCZAK J M. Manifold regularized particle filter for articulated human motion tracking [C]// ICSS 2013: Proceedings of the International Conference on Systems Science. Berlin: Springer, 2014: 283–293.

(下转第 2326 页)

- [12] LOPEZ R, HUANG T S. 3D Head pose computation from 2D images: template versus features [C]// ICIP '95: Proceedings of the 1995 International Conference on Image Processing. Washington, DC: IEEE Computer Society, 1995, 2: 599 - 602.
- [13] YILMAZ A, SHAH M. Automatic feature detection and pose recovery of faces [C]// ACCV 2002: Proceedings of the 5th Asian Conference on Computer Vision. Tokyo: Asian Federation of Computer Vision Societies, 2002: 23 - 35.
- [14] YANG R, ZHANG Z. Model-based head pose tracking with stereo vision [C]// Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition. Piscataway: IEEE, 2002: 255 - 260.
- [15] DEMENTHON D F, DAVIS L S. Model-based object pose in 25 lines of code [J]. International Journal of Computer Vision, 1995, 15(1/2): 123 - 141.
- [16] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF [C]// ICCV '11: Proceedings of the 2011 International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2011: 2564 - 257.
- [17] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [M]// Readings in Computer Vision: Issues, Problems, Principles, and Paradigms. San Francisco: Morgan Kaufmann Publishers, 1987: 726 - 740.
- [18] CAO X. Face alignment by explicit shape regression [C]// CVPR 2012: Proceedings of the 1996 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2012: 2887 - 2894.
- [19] CAO C, HOU Q, ZHOU K. Displaced dynamic expression regression for real-time facial tracking and animation [J]. ACM Transactions on Graphics: Proceedings of ACM SIGGRAPH 2014, 2014, 33(4): Article No. 43.
- [20] CAO C, WENG Y, LIN S, et al. 3D shape regression for real-time facial animation [J]. ACM Transactions on Graphics: SIGGRAPH 2013 Conference Proceedings, 2013, 32(4): Article No. 41.
- [21] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection [C]// Proceedings of the 2006 European Conference on Computer Vision. Berlin: Springer, 2006: 430 - 443.
- [22] CALONDER M, LEPEITI V, STRECHA C, et al. BRIEF: Binary Robust Independent Elementary Features [C]// ECCV 2010: Proceedings of the 11th European Conference on Computer Vision, LNCS 6314. Berlin: Springer, 2010: 778 - 792.
- [23] LOWE D G. Distinctive image features from scale invariant features [J]. International Journal of Computer Vision, 2004, 60(2): 91 - 110.
- [24] BAY H, ESS A, TUYTELAARS T, et al. SURF: Speeded Up Robust Features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346 - 359.
- [25] BYRÖD M, ÅSTRÖM K. Bundle adjustment using conjugate gradients with multiscale preconditioning [C]// BMVC 2009: Proceedings of the 2009 British Machine Vision Conference. London: Lund University Publications, 2009: 29 - 42.
- [26] LOURAKIS M I A, ARGYROS A A. SBA: a software package for generic sparse bundle adjustment [J]. ACM Transactions on Mathematical Software, 2009, 36(1): Article No. 2.
- [27] MEI P. Research on real-time face detection and pose estimation in video sequence [D]. Wuhan: Wuhan University of Technology, 2011: 46 - 48. (梅鹏. 视频序列中实时人脸检测及姿态估计的研究[D]. 武汉: 武汉理工大学, 2011: 46 - 48.)
- [28] LIANG G, ZHA H, LIU H. Face pose estimation based on 3D models and affine correspondences [J]. Chinese Journal of Computers, 2005, 28(5): 792 - 800. (梁国远, 查红彬, 刘宏. 基于三维模型和仿射对应原理的人脸姿态估计方法[J]. 计算机学报, 2005, 28(5): 792 - 800.)

(上接第 2315 页)

- [10] LUBER M, STORK J A, TIPALDI G D, et al. People tracking with human motion predictions from social forces [C]// Proceedings of the 2010 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2010: 464 - 469.
- [11] WU B, NEVATIA R. Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors [J]. International Journal of Computer Vision, 2007, 75(2): 247 - 266.
- [12] MILAN A, SCHINDLER K, ROTH S. Detection-and trajectory-level exclusion in multiple object tracking [C]// CVPR 2013: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2013: 3682 - 3689.
- [13] LI H, ZHAO R, WANG X. Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification [J/OL]. arXiv preprint arXiv: 1412.4526, 2014 [2014-12-16]. <http://arxiv.org/abs/1412.4526>.
- [14] YANG B, NEVATIA R. Multi-target tracking by online learning a CRF model of appearance and motion patterns [J]. International Journal of Computer Vision, 2014, 107(2): 203 - 217.
- [15] HELBING D, MOLNÁR P. Social force model for pedestrian dynamics [J]. Physical Review E, 1995, 51(5): 4282 - 4286.
- [16] PELLEGRINI S, ESS A, SCHINDLER K, et al. You'll never walk alone: modeling social behavior for multi-target tracking [C]// Proceedings of the 2009 IEEE 12th International Conference on Computer Vision. Piscataway: IEEE, 2009: 261 - 268.
- [17] ANDRIYENKO A, SCHINDLER K, ROTH S. Discrete-continuous optimization for multi-target tracking [C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2012: 1926 - 1933.
- [18] FELZENZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627 - 1645.
- [19] WANG X, MA X, GRIMSON W E L. Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(3): 539 - 555.
- [20] LI Y, HUANG C, NEVATIA R. Learning to associate: Hybrid-boosted multi-target tracker for crowded scene [C]// CVPR 2009: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2009: 2953 - 2960.
- [21] STIEFELHAGEN R, BERNARDIN K, BOWERS R, et al. The CLEAR 2006 evaluation [C]// CLEAR '06: Proceedings of the 1st International Evaluation Conference on Classification of Events, Activities and Relationships. Berlin: Springer, 2007: 1 - 44.